Quickly forecasting the future state of urban sensors by the missing-data-tolerant deep learning approach

Peixiao Wang [a,b,c], Hengcai Zhang [a,b,*], Shifen Cheng [a,b], Tong Zhang [c], Feng Lu [a,b,d]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, 100101, China
[b] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, China
[c] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[d] Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou, 350003, China

ARTICLE INFO

ABSTRACT

Accurately and quickly forecasting the future state of urban sensors is crucial for urban monitoring and management. Although many forecasting approaches have been proposed, existing models still face two major challenges. First, most approaches do not have the ability to automatically handle missing data. Second, most approaches have high complexity, neglecting the usability and lightweight of the approach. Therefore, we present a lightweight spatiotemporal dilation approach tolerating missing data (STDM) to address the aforementioned challenges. First, we integrate a missing data handling mechanism into the STDM approach to enhance its forecasting capability under missing scenarios. Second, we present a lightweight spatiotemporal dilation component to enhance the inference speed of the STDM approach. Finally, we design the STDM approach as a separable architecture and define a corresponding loss function, allowing the STDM approach to be compatible with both forecasting tasks under missing and non-missing scenarios. The approach underwent validation using traffic, PM$_{2.5}$, and temperature datasets. It exhibited superior forecasting accuracy and inference speed across four missing scenarios, outperforming eight baselines. Codes and data are available at link on https://doi.org/10.6084/m9.figshare.24289456.

1. Introduction

In the age of the Internet of Things, sensor networks have expanded significantly, allowing for fine-grained monitoring of the state of cities (Zhao et al., 2019; Yu, 2021; Z. Xu et al., 2023). With the explosion of sensor data in urban, how to use sensor data to forecast the future state of cities has attracted attention (Jiang et al., 2022; Liu et al., 2023; Zheng et al., 2014).

As depicted in Fig. 1, the core of urban state forecasting involves establishing a mapping function between input and output. This function is then utilized to extrapolate unknown data to forecast time points (Khaled et al., 2024; Liu et al., 2022; Qu et al., 2023). Building upon these modeling concepts, various approaches are proposed for urban state forecasting (P. Wang et al., 2024; Z. Wang et al., 2024; B. Zhang et al., 2023). However, most forecasting approaches still encounter two significant challenges:

(1) **Lack of ability to automatically handle missing data**: Most existing forecasting approaches are unable to handle missing data, as they are typically tested on datasets that assume no data is missing (Chen & Sun, 2022; Cui et al., 2020). In real-world scenarios, however, missing data is frequent and often exhibits complex patterns, such as random or block missing (Li et al., 2020). This limitation reduces the applicability and accuracy of many forecasting models.

(2) **Neglecting the usability and lightweight of the approach**: Although most existing forecasting approaches improve the forecasting accuracy, they also suffer from overly complex models (M. Xu et al., 2021; Zhang et al., 2020). In real scenarios, it is crucial to quickly forecast the future state of the city. However, excessive model complexity limits the forecasting speed of the existing approaches (Do et al., 2019; Li et al., 2023).

* Corresponding author.
E-mail addresses: wpx@lreis.ac.cn (P. Wang), zhanghc@lreis.ac.cn (H. Zhang), chengsf@lreis.ac.cn (S. Cheng), zhangt@whu.edu.cn (T. Zhang), luf@lreis.ac.cn (F. Lu).

Overall, we urgently need to develop new urban state forecasting approaches that can automatically handle missing data and are efficient and user-friendly. Therefore, this study presented a lightweight spatio-temporal dilation approach tolerating missing data (STDM), with the following contributions:

(1) This study introduced a new Missing Data Handling component (MDH) to enable STDM approach to automatically manage missing values based on their patterns. Additionally, this study presented a lightweight spatiotemporal dilation component (STD) that allows the STDM approach to quickly infer the future state of urban sensors.

(2) To ensure the adaptability of the STDM approach across spatio-temporal forecasting tasks, whether data is incomplete or complete, this study structured it with an Imputer-Predictor architecture. Moreover, this study designed a corresponding loss to train learnable weights by addressing both imputation and forecasting tasks concurrently.

(3) The STDM approach underwent validation using traffic, PM$_{2.5}$, and temperature datasets, assessing its forecasting accuracy and forecasting speed, especially in missing scenes. Additionally, this study has open-sourced the code for the STDM approach to ensure reproducibility.

2. Literature review

Urban sensor state forecasting is essentially spatio-temporal forecasting. This section mainly reviewed the current research status from the perspective of whether spatiotemporal forecasting approaches tolerate missing data.

2.1. Spatiotemporal forecasting approaches without tolerating missing data

Missing-data-intolerant spatiotemporal forecasting approaches are currently the mainstream approaches, which build the function mapping relationship between input data and output data based on complete spatiotemporal data (or spatiotemporal data without missing data) (L. Xu et al., 2021). The missing-data-intolerant spatiotemporal forecasting approaches are often modeled in multiple stages to handle missing data in the spatiotemporal data. Specifically, relevant scholars first estimated missing data in spatiotemporal data (or deleted spatiotemporal data containing missing data), and then established forecasting approaches based on the processed spatiotemporal data (Du et al., 2021; Fang et al., 2022; Liu et al., 2022). Common missing-data-intolerant forecasting approaches include the ST-KNN approach (Cheng et al., 2018), the ST-GCN approach (Diao et al., 2019), the ST-ResNet approach (Zhang et al., 2017), the T-GCN approach (Zhao et al., 2020), the ASTGCN approach (Guo et al., 2019), and the GDGCN approach (Y. Xu et al., 2023). At present, missing-data-intolerant forecasting approaches have achieved high forecasting accuracy in spatiotemporal datasets without missing data, but there are still shortcomings. For example, the above approaches accomplish the forecasting tasks under missing scenarios with the help of preprocessing operations (introducing imputation models or deleting partial spatiotemporal data) (Cheng et al., 2018). The former increases computational burden, with imputation accuracy directly impacting forecasting approach performance, while the latter may result in insufficient training data, failing to capture reliable spatiotemporal patterns (P. Wang, Zhang, Hu et al., 2023).

2.2. Spatiotemporal forecasting approaches with tolerating missing data

In contrast to approaches intolerant to missing data, those tolerant to missing data directly utilize raw incomplete data (or spatiotemporal data with missing values) to build mapping function between input and output (Mei et al., 2023). Early missing-data-tolerant forecasting approaches, like the GRU-D approach (Che et al., 2018) and the LSTM-M approach (Tian et al., 2018), suffer from limited forecasting accuracy because of the challenge in uncovering the complexity of spatial relationships. In recent years, to improve the forecasting accuracy under missing scenarios, some scholars further proposed forecasting models that simultaneously depict spatial and temporal information, including the TRMF approach (Yu et al., 2016), the BTMF approach (Chen & Sun, 2022), the Causal-GCNM approach (P. Wang, Zhang, Nie et al., 2023), and the D-TGNM approach (P. Wang, Zhang, Hu et al., 2023). In general, compared to missing-data-intolerant forecasting approaches, the missing-data-tolerant forecasting approaches are very limited. In addition, most of the above forecasting approaches suffer from overly complex models, affecting the inference speed of forecasting approaches (Li et al., 2023; P. Wang et al., 2024). In real scenarios, the inference speed of the approach is crucial. However, the existing approaches are difficult to balance inference speed and forecasting accuracy.

To address the shortcomings of the above approaches, this study presented a new forecasting approach for fast forecasting of urban sensor states in missing scenes. Compared with the existing approaches, the proposed STDM approach can not only directly complete forecasting tasks in missing scenarios, but also has a faster inference speed.
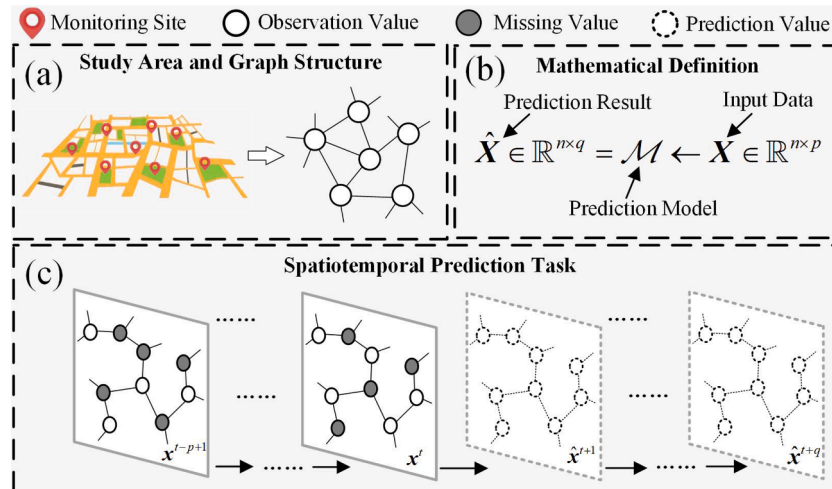

Fig. 1. Basic process of urban state forecasting.

3. Methodology

3.1. Preliminaries

Considering the generality of the graph structure, this study constructed the STDM method based on the graph structure. In this section, this study first provided the necessary definitions for the STDM approach and describe the mathematical formulation of the urban sensor state forecasting.

Definition 1. The sensors within the study area collectively form a graph structure, denoted as $G = \left\langle \{v_i\}_{i=1}^n, \ \boldsymbol{A} \in \mathscr{R}^{n \times n}, \ \boldsymbol{X} \in \mathscr{R}^{n \times T} \right\rangle$, where v_i indicates the ith sensor, with n being the total number of sensors; $\boldsymbol{A}$ indicates the adjacency matrix between sensors; $\boldsymbol{X} \in \mathscr{R}^{n \times T}$ signifies the spatiotemporal data of n sensors across T time windows. Additionally, this study defined $\boldsymbol{x}^t = \{x_i^t\}_{i=1}^n \in \boldsymbol{X}$ as the space vector of n sensors in the tth time window, with $x_i^t \in \boldsymbol{X}$ representing the data of ith sensor.

Definition 2. An indicator matrix $\boldsymbol{M} \in \mathscr{R}^{n \times T}$ is utilized to differentiate missing data from non-missing data in matrix $\boldsymbol{X} \in \mathscr{R}^{n \times T}$. If $m_i^t = 0$, the spatiotemporal data x_i^t is missing, and if $m_i^t = 1$, the spatiotemporal data x_i^t is not missing. Similarly, $\boldsymbol{m}^t = \{m_i^t\}_{i=1}^n \in \mathscr{R}^{n \times 1}$ is utilized to differentiate missing data from non-missing data in vector $\boldsymbol{x}^t = \{x_i^t\}_{i=1}^n \in \mathscr{R}^{n \times 1}$.

This study aims to develop a novel forecasting approach, which can model spatiotemporal data with missing values and quickly infer future spatiotemporal data, as depicted in Eq. (1)

$$\widehat{\boldsymbol{X}}_{t+1}^{t+q} = \left\{\widehat{\boldsymbol{x}}^{t+1}, \widehat{\boldsymbol{x}}^{t+2}, \ldots\ldots, \widehat{\boldsymbol{x}}^{t+q}\right\} = \mathscr{M} \ \leftarrow \ \left\langle \boldsymbol{X}_{t-p+1}^t, \boldsymbol{M}_{t-p+1}^t; \boldsymbol{W} \right\rangle \quad (1)$$

where $\mathscr{M}$ signifies the proposed STDM approach; $\boldsymbol{X}_{t-p+1}^t = \{\boldsymbol{x}^\tau\}_{\tau=t-p+1}^t$ represents the historical data, with p being the time-dependent step; $\boldsymbol{M}_{t-p+1}^t = \{\boldsymbol{m}^\tau\}_{\tau=t-p+1}^t$ indicates the indicator matrix of $\boldsymbol{X}_{t-p+1}^t$; $\widehat{\boldsymbol{x}}^{t+q}$ indicates the future data, with q being the forecasting step; $\boldsymbol{W}$ signifies the learnable parameters.

3.2. Proposed STDM approach

3.2.1. Architecture and forward propagation of the STDM

Fig. 2 depicts the STDM approach, presenting an Imputer-Predictor architecture that can be employed for two prediction tasks in both missing and non-missing scenarios. In non-missing scenarios, the prediction task can be completed using only the Predictor. In contrast, for missing scenarios, it is only necessary to add the Imputer to the Predictor. This design enables the STDM to effectively adapt to both missing and non-missing scenarios. Moreover, the Imputer and Predictor are structurally identical, differing only in their loss function definitions and inputs. On the left, a MDH component autonomously handles missing data based on missing patterns (as discussed in Section 3.2.2). On the right, multiple STD components extract spatiotemporal correlations from data and quickly infer the future state of urban sensors (as discussed in Section 3.2.3).

Supplementary Figure S1 provides a detailed depiction of the internal structure of the STDM approach. Typically, the data gathered by sensors are fed into Imputer, producing a temporary variable. Subsequently, this variable serves as input for Predictor, culminating in the ultimate forecasting. Since Imputer and Predictor have identical internal structures, their input data processing is analogous, involving one MDH component and N_I/N_P STD components, as shown in Eqs. (2) and (3).
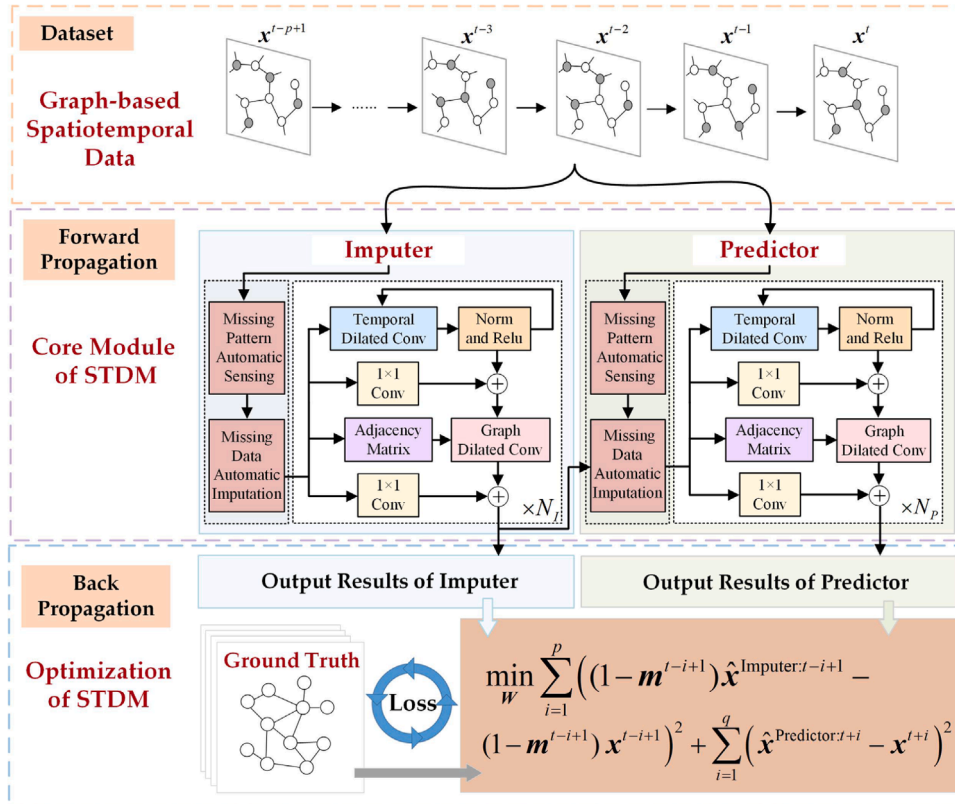

Fig. 2. The overarching framework of the proposed approach.

$$\text{Imputer}: \begin{cases} \overline{\boldsymbol{X}}_{t-p+1}^{t} = MDH\left(\boldsymbol{X}_{t-p+1}^{t}, \boldsymbol{X}_{t-p+1}^{t}, \boldsymbol{M}_{t-p+1}^{t}; \boldsymbol{W}\right) \\ \mathscr{O}_{l:t-p+1}^{l:t} = \begin{cases} STD\left(\overline{\boldsymbol{X}}_{t-p+1}^{t}; \boldsymbol{W}\right) & l=1 \\ STD\left(\mathscr{O}_{l-1:t-p+1}^{l-1:t}; \boldsymbol{W}\right) & l < N_I \end{cases} \\ \left\{\widehat{\boldsymbol{x}}^{\text{Imputer}:\tau}\right\}_{\tau=t-p+1}^{t} = Conv\left(\mathscr{O}_{N_I:t-p+1}^{N_I:t}\right) \end{cases} \quad (2)$$

$$\text{Predictor}: \begin{cases} \overline{\boldsymbol{X}}_{t-p+1}^{t} = MDH\left(\boldsymbol{X}_{t-p+1}^{t}, \left\{\widehat{\boldsymbol{x}}^{\text{Imputer}:\tau}\right\}_{\tau=t-p+1}^{t}, \boldsymbol{M}_{t-p+1}^{t}; \boldsymbol{W}\right) \\ \mathscr{O}_{l:t-p+1}^{l:t} = \begin{cases} STD\left(\overline{\boldsymbol{X}}_{t-p+1}^{t}; \boldsymbol{W}\right) & l=1 \\ STD\left(\mathscr{O}_{l-1:t-p+1}^{l-1:t}; \boldsymbol{W}\right) & l < N_P \end{cases} \\ \left\{\widehat{\boldsymbol{x}}^{\text{Predictor}:\tau}\right\}_{\tau=t+1}^{t+q} = Conv\left(\mathscr{O}_{N_P:t-p+1}^{N_P:t}\right) \end{cases} \quad (3)$$

where $\boldsymbol{X}_{t-p+1}^{t} \in \mathscr{R}^{n \times p}$, $\boldsymbol{M}_{t-p+1}^{t} \in \mathscr{R}^{n \times p}$, p, q, and $\boldsymbol{W}$ represent the same meaning as in Equation (1); $\left\{\widehat{\boldsymbol{x}}^{\text{Predictor}:\tau}\right\}_{\tau=t+1}^{t+q} \in \mathscr{R}^{n \times q}$ denotes the ultimate forecasting result of the STDM approach, equivalent to $\widehat{\boldsymbol{X}}_{t+1}^{t+q}$ in Equation (1); $\left\{\widehat{\boldsymbol{x}}^{\text{Imputer}:\tau}\right\}_{\tau=t-p+1}^{t} \in \mathscr{R}^{n \times p}$ denotes the outcome of the Imputer; $\overline{\boldsymbol{X}}_{t-p+1}^{t} = \{\overline{\boldsymbol{x}}^{\tau}\}_{\tau=t-p+1}^{t} \in \mathscr{R}^{n \times p}$ signifies the outcome of the MDH; $\mathscr{O}_{l:t-p+1}^{l:t} \in \mathscr{R}^{n \times p \times e}$ signifies the outcome of the lth STD, with e being the data dimension; $Conv$ signifies the convolution function for dimension alignment. Eqs. (2) and (3) illustrate that the primary distinction between Imputer and Predictor is their respective inputs and outcomes. Additionally, these equations highlight that the MDH and STD components are central to both the Imputer and Predictor.

3.2.2. Missing data handling component of the STDM

The crux of the MDH component lies in identifying missing patterns accurately. Therefore, this study defined an auxiliary quantity $\boldsymbol{u}^{t} = \left\{u_i^t\right\}_{i=1}^{n}$ to identify different types of missing patterns, with u_i^t being the time step of x_i^t from the nearest observation (positive or negative directions of the timeline). Note: this study considers the minimum value of u_i^t in both directions.

Take x_i^{t-1} as an example (red node), Fig. 3 illustrates the calculation process for u_i^{t-1}. If the data x_i^{t-1} is missing, u_i^{t-1} equals 2 along the forward direction of the timeline and 1 along the reverse direction. Taking the minimum in both directions results in u_i^{t-1} being equal to 1. When u_i^{t-1} is obtained, we can judge the type of missing pattern of x_i^{t-1} by the u_i^{t-1}. For instance, if $u_i^{t-1} = 1$, it suggests that x_i^{t-1} tends to be random missing, whereas $u_i^{t-1} > 1$ indicates that x_i^{t-1} tends to be block missing. As an example, Eq. (4) demonstrates the computation process of the $\boldsymbol{u}^t = \left\{u_i^t\right\}_{i=1}^{n}$ in the forward direction along the time axis.

$$u_i^t = \begin{cases} 1 + u_i^{t-1} & t > 1, m_i^{t-1} = 0 \\ 1 & t > 1, m_i^{t-1} = 1 \\ 0 & t = 1 \end{cases} \quad (4)$$

where u_i^t and m_i^{t-1} represent the same meaning as before.

Upon $\boldsymbol{u}^t$ is obtained, we extend the functionality of the MDH component to autonomously address missing data. Broadly speaking, when confronted with random missing patterns, the correlation between missing data and temporal observations tends to outweigh that with spatial observations. However, as the pattern shifts towards block missing, this correlation with temporal observations gradually wanes, while the connection with spatial observations strengthens (P. Wang, Zhang, Hu et al., 2023). This adaptive behavior is crucial for effectively handling missing data across diverse patterns, ensuring robust estimations within our model framework. Drawing inspiration from these concepts, we adopt a methodology where the estimation of missing data involves a weighted aggregation of the spatially most similar observation and the temporally closest observation, as shown in Eq. (5).

$$\text{MDH}: \begin{cases} \overline{\boldsymbol{x}}^t = \boldsymbol{m}^t \odot \boldsymbol{x}^t + (1 - \boldsymbol{m}^t) \odot (\boldsymbol{\beta}^t \odot \boldsymbol{x}^{t:tm} + (1 - \boldsymbol{\beta}^t) \odot \boldsymbol{x}^{t:sm}) \\ \boldsymbol{\beta}^t = \exp\{ - \max(0, \boldsymbol{W}_\delta \boldsymbol{u}^t + \boldsymbol{b}_\delta)\} \end{cases} \quad (5)$$

where $\overline{\boldsymbol{x}}^t$ signifies processed spatiotemporal data through the MDH, and the processed data within p time windows is encapsulated by $\overline{\boldsymbol{X}}_{t-p+1}^{t} = \{\overline{\boldsymbol{x}}^{\tau}\}_{\tau=t-p+1}^{t} \in \mathscr{R}^{n \times p}$; $\boldsymbol{x}^{t:tm} \in \mathscr{R}^{n \times 1}$ signifies the nearest data of target nodes in the time dimension; $\boldsymbol{x}^{t:sm} \in \mathscr{R}^{n \times 1}$ represents the most akin data of target nodes in the spatial dimension; $\boldsymbol{\beta}^t$ represents the probability of the missing pattern type calculated by $\boldsymbol{u}^t$ for automatic processing of missing data. **Note:** Eq. (5) solely delineates the forward propagation process of the MDH component and does not really impute the missing data. The imputation strategy of the MDH component will be learned automatically during model optimization.

3.2.3. Spatiotemporal dilation component of the STDM

Following processing by the MDH component, the STD component extracts spatiotemporal correlations from data and quickly infer the future state of urban sensors. Illustrated in Figure 4(b), the STD component primarily includes temporal dilation convolution operator (TDC) and graph dilation convolution operator (GDC). Compared to the normal operator, the TDC operator and GDC operator enhance the computational efficiency by reducing the number of forward propagation (P. Wang et al., 2024). Eq. (6) explicitly details the forward propagation mechanism inherent to the T-GA component.
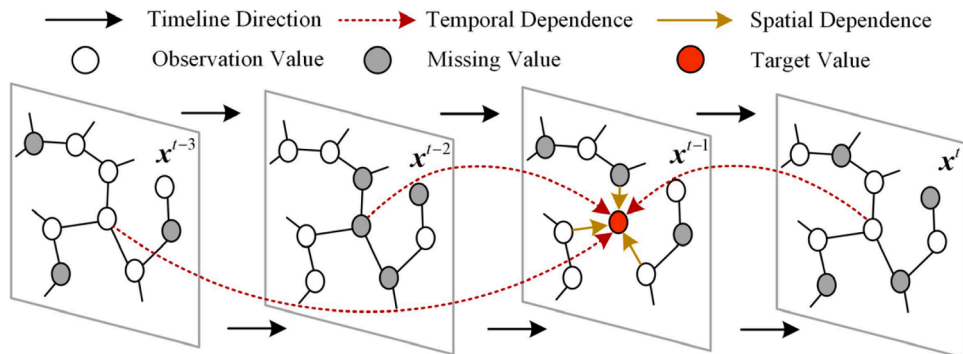

Fig. 3. llustration of missing data handling component.

$$STD : \begin{cases} \mathscr{H} = TDC_d\left(\overline{\boldsymbol{X}}^t_{t-p+1}\right) \\ \mathscr{H}' = Relu(Norm(\mathscr{H})) \\ \mathscr{H}'' = Relu(Norm(TDC_d(\mathscr{H}'))) + Conv\left(\overline{\boldsymbol{X}}^t_{t-p+1}\right) \\ \mathscr{O} = GDC_d(\mathscr{H}'') \end{cases} \quad (6)$$

where $\mathscr{H} \in \mathscr{R}^{n \times p \times e}$, $\mathscr{H}' \in \mathscr{R}^{n \times p \times e}$, and $\mathscr{H}'' \in \mathscr{R}^{n \times p \times e}$ represent the temporary states; Taking $\mathscr{H}$ as an example, it can be decomposed into $\{\boldsymbol{H}_i\}_{i=1}^n$ along the space dimension and $\{\boldsymbol{H}^\tau\}_{\tau=t-p+1}^t$ along the time dimension; Here, $\{\boldsymbol{H}_i\}_{i=1}^n$ is utilized for modeling of space dimension, while $\{\boldsymbol{H}^\tau\}_{\tau=t-p+1}^t$ is employed for modeling of time dimension; $\mathscr{O} \in \mathscr{R}^{n \times p \times e}$ denotes the output of STD Component; $Conv\left(\overline{\boldsymbol{X}}^t_{t-p+1}\right)$ represents performing convolution operations on $\overline{\boldsymbol{X}}^t_{t-p+1}$.

Figs. 4(a) and (c) further illustrate the computational details of the TDC and GDC operators. In the TDC operator, the dilatation factor is used to expand the receptive field of the convolution operation along the time dimension to improve the operational efficiency of the proposed approach. Similarly, the GDC operator utilizes the dilatation factor to improve the computational efficiency. In addition, in order to improve the forecasting accuracy of the proposed approach, we define the forward propagation of the GDC operator based on graph attention. More specifically, Eqs. (7) and (8) further show the calculation process of the TDC operator and the GA operator.

$$TDC_d\left(\overline{\boldsymbol{X}}^t_{t-p+1}\right) = \begin{pmatrix} \sum_{l=1}^L w_l\overline{x}_n^{(t-p-1)-(L-l)d} & \cdots & \sum_{l=1}^L w_l\overline{x}_n^{t-(L-l)d} \\ \vdots & \ddots & \vdots \\ \sum_{l=1}^L w_l\overline{x}_1^{(t-p-1)-(L-l)d} & \cdots & \sum_{l=1}^L w_l\overline{x}_1^{t-(L-l)d} \end{pmatrix} \quad (7)$$

$$GDC_d(\mathscr{H}'') = \begin{cases} \boldsymbol{O}_i = \sum_{j \in \boldsymbol{A}_i^d} \gamma_{ji}\boldsymbol{H}_i''\boldsymbol{W}_v \\ \gamma_{ji} = \dfrac{\exp\left(Relu\left(\left[\boldsymbol{H}_i'' \middle\| \boldsymbol{H}_j''\right]\boldsymbol{W}_q\right)\right)}{\sum_{k \in \boldsymbol{A}_i^d}\exp\left(Relu\left(\left[\boldsymbol{H}_i'' \middle\| \boldsymbol{H}_k''\right]\boldsymbol{W}_q\right)\right)} \end{cases} \quad (8)$$

where $TDC_d\left(\overline{\boldsymbol{X}}^t_{t-p+1}\right)$ denotes the temporal dilation convolution operation on $\overline{\boldsymbol{X}}^t_{t-p+1}$ with dilation factor d; $\{w_l\}_{l=1}^L$ denotes the convolution kernel in the TDC operator, with L being kernel size; For dimensional alignment during operations, the TDC tends to contain e convolution kernel; $GDC_d(\mathscr{H}'')$ denotes the graph dilation convolution operation on $\mathscr{H}''$ with dilation factor d; γ_{ji} indicates the influence weight of graph node v_j on graph node v_i; The meaning of $\boldsymbol{H}_i''$ is the same as that of $\boldsymbol{H}_i''$ in Equation (6); $\boldsymbol{A}_i^d$ represents the dilation adjacency matrix, calculated using the method provided by P. Wang et al. (2024); $\boldsymbol{W}_q$ and $\boldsymbol{W}_v$ represent learnable parameters in the GDC; $Relu$ stands for activation function; $[\bullet\|\bullet]$ represents the matrix join function.

3.2.4. Loss function of the STDM

In the previous context, the imputation result $\{\hat{\boldsymbol{x}}^{\text{Imputer}:\tau}\}_{\tau=t-p+1}^t$ for p time windows is first obtained through the Imputer, and then the spatiotemporal data $\{\hat{\boldsymbol{x}}^{\text{Predictor}:\tau}\}_{\tau=t+1}^{t+q}$ for future q time windows is obtained through the Predictor. In the general forecasting task, the final STDM approach can be obtained by optimizing the square loss between the truth value $\{\boldsymbol{x}^\tau\}_{\tau=t+1}^{t+q}$ and the forecasting value $\{\hat{\boldsymbol{x}}^{\text{Predictor}:\tau}\}_{\tau=t+1}^{t+q}$. However, in this study, optimizing only the square loss between $\{\boldsymbol{x}^\tau\}_{\tau=t+1}^{t+q}$ and $\{\hat{\boldsymbol{x}}^{\text{Predictor}:\tau}\}_{\tau=t+1}^{t+q}$ often fails to obtain better forecasting accuracy. The reason is that the above loss ignores the effect of the Imputer (or missing data) on the forecasting results. Therefore, this study designed a new loss to optimize learnable parameters.

In Fig. 5, the designed loss function primarily includes the loss of the Predictor, ensuring the accuracy of forecasting results, and the loss of the Imputer, accounting for missing data's impact on forecasting results. Eq. (9) further shows this function, with additional details on model training provided in Appendix A.

$$\mathscr{L}(\boldsymbol{W}) = \min_{\boldsymbol{W}} \left(\begin{array}{l} \sum_{i=1}^p \left(\left(1 - \boldsymbol{m}^{t-i+1}\right)\hat{\boldsymbol{x}}^{\text{Imputer}:t-i+1} - \left(1 - \boldsymbol{m}^{t-i+1}\right)\boldsymbol{x}^{t-i+1}\right)^2 \\ + \sum_{i=1}^q \left(\hat{\boldsymbol{x}}^{\text{Predictor}:t+i} - \boldsymbol{x}^{t+i}\right)^2 \end{array} \right) \quad (9)$$

where $\sum_{i=1}^p \left(\left(1 - \boldsymbol{m}^{t-i+1}\right)\hat{\boldsymbol{x}}^{\text{Imputer}:t-i+1} - \left(1 - \boldsymbol{m}^{t-i+1}\right)\boldsymbol{x}^{t-i+1}\right)^2$ represents the loss of Imputer; $\sum_{i=1}^q \left(\hat{\boldsymbol{x}}^{\text{Predictor}:t+i} - \boldsymbol{x}^{t+i}\right)^2$ represents the loss of Predictor.

4. Experiment

4.1. Data sources and preprocessing

The traffic, PM2.5, and temperature datasets are widely used urban sensor datasets. Numerous studies currently utilize these three types of data to investigate traffic conditions, air quality, and climate change in urban areas(Palanisamy et al., 2024; Rabie et al., 2024; Soudeep et al., 2024; M. Zhang et al., 2023). Therefore, we utilize these datasets to validate the performance of the STDM approach. Fig. 6 illustrates the
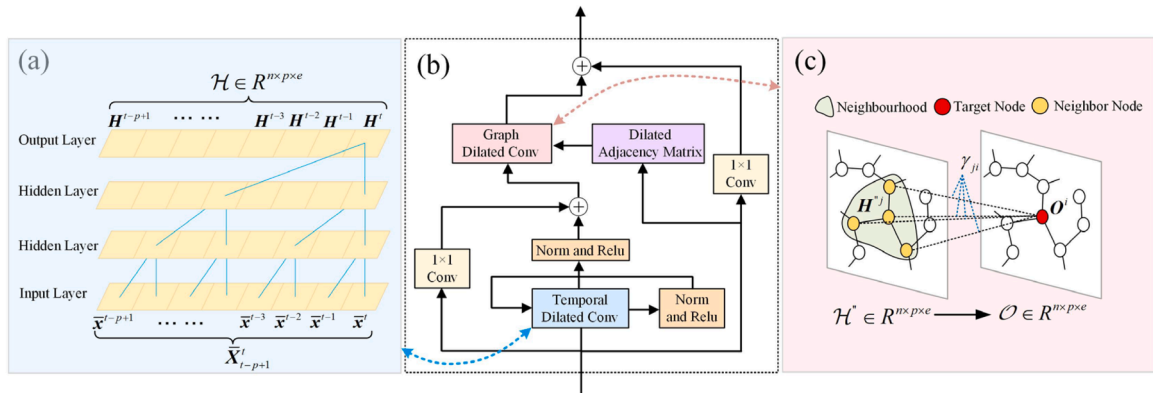

Fig. 4. llustration of spatiotemporal dilation component: (a) temporal dilatation convolution operator, (b) forward propagation of the STD component, and (c) graph dilatation convolution operator.
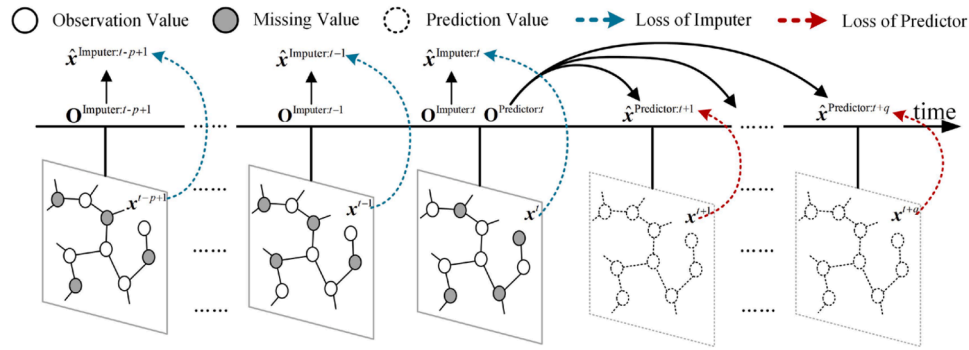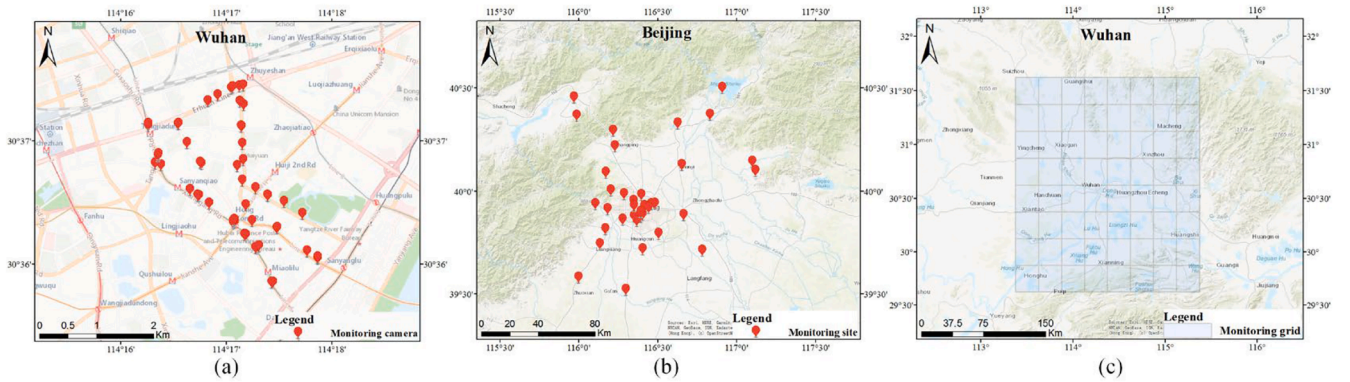
Fig. 5. Loss function illustration of the STDM.

Fig. 6. Spatial distribution of datasets: (a) traffic dataset, (b) PM$_{2.5}$ dataset, and (c) temperature dataset.

spatial distribution of these three datasets, while Table 1 presents their statistical characteristics. In addition, we mainly preprocess the raw data as follows.

(1) Natural Missing Data Estimation: The raw data contained natural missing data. In this study, we first utilized the BTTF model to estimate them (Chen & Sun, 2022).
(2) Missing Data Generation: Following the methods outlined by Wang et al. (2023) and Cui et al. (2020), we introduced two types of missing data—random missing and block missing—by removing some spatiotemporal data with missing rates of 20 % and 40 %. For instance, in the PM$_{2.5}$ dataset, Fig. 7 illustrates the processed spatiotemporal data information.
(3) Adjacency Matrix Construction: We established adjacency relationships for the three datasets. In this process, the adjacency relationships were derived from a similarity matrix, wherein the ten most similar spatial objects were designated as neighbors for each target spatial object.

Table 1
Statistical characteristics of the datasets.

Dataset	Traffic	PM$_{2.5}$	Temperature
Spatial distribution	Fig. 6(a)	Fig. 6(b)	Fig. 6(c)
Indicator	Traffic Volume	PM$_{2.5}$ Concentration	Air Temperature
Spatial objects	67	36	64
Temporal objects	8064	2952	2208
Natural missing rate	8.94 %	10.35 %	0 %
Time window	5 min	60 min	60 min
Time span	2021/3/1~2021/3/28	2014/5/1~2014/8/31	2018/6/1~2018/8/31

4.2. Calibration of hyper-parameters

This subsection outlines the experimental settings, including the hardware setting, software setting, and hyper-parameter setting.

The processing of spatiotemporal data took place on a PC featuring an Intel(R) Core(TM) i7-11,700 CPU @ 2.50 GHz and 16GB of memory. Furthermore, our model was developed using PyTorch and Python 3.7, and executed on a GPU platform with 24GB of GPU memory.

The hyper-parameters of the STDM approach primarily include the time-dependent step p, hidden state dimension e, kernel size K, dilation factor d, number of T-GA in Imputer N_I, and number of STD in Predictor N_P. Since the STDM approach is an Imputer-Predictor architecture, we can first calibrate the hyper-parameters in the Imputer, and then calibrate the hyper-parameters in the Predictor. In this study, we first set the time-dependent step to 10, the kernel size to 3, and the dilation factor to 2. Then, we calibrated the hidden state dimension e, the number N_I of STD in Imputer, and the number N_P of STD in Predictor. Fig. 8 illustrates the calibration process of the hyper-parameters on complete traffic dataset. The results indicate that model prediction accuracy gradually improves and stabilizes as the hyperparameters increase. Based on these findings, the hidden state dimension was set to 32, the number of STD in Imputer to 2, and the number of STD in Predictor to 2.

4.3. Quantitative analysis of forecasting accuracy

4.3.1. Selection of baseline methods

In this study, we employ RMSE and MAPE as metrics to evaluate the accuracy of the proposed approach. Given that classical statistical approaches often exhibit lower performance on spatiotemporal forecasting tasks compared to data-driven approaches, our primary comparison focuses on evaluating the STDM approach against popular data-driven approaches, falling into two categories:
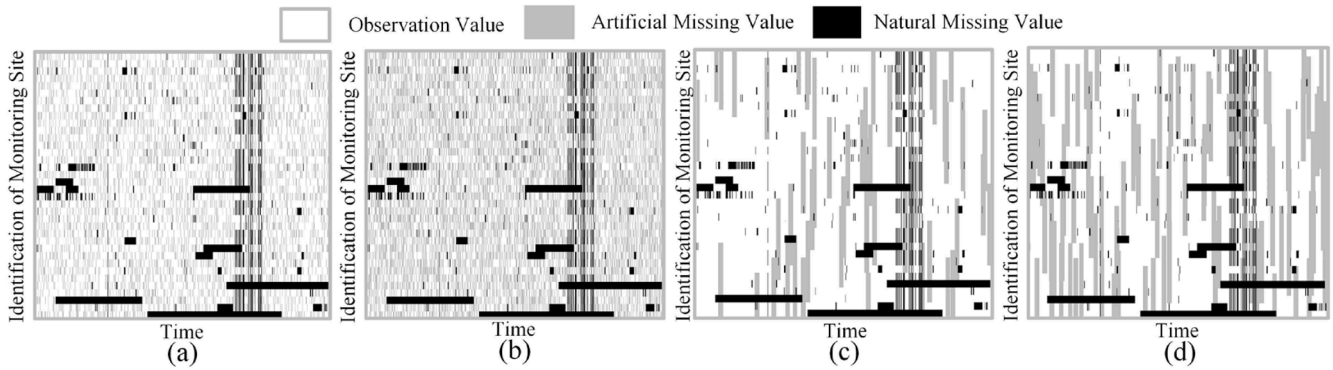
Fig. 7. Location distribution of missing data in the PM$_{2.5}$ dataset: (a) 20 % missing rate with random missing, (b) 40 % missing rate with random missing, (c) 20 % missing rate with block missing, and (d) 40 % missing rate with block missing.
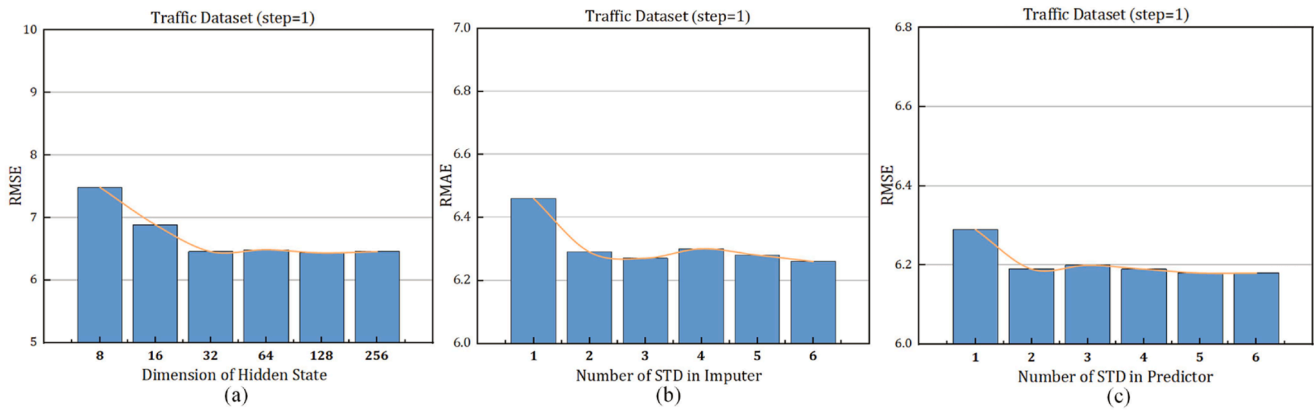
Fig. 8. Hyper-parameter tuning of the STDM approach: (a) hidden state dimension,(b) number of STD in Imputer, and (c) number of STD in Predictor.

- Forecasting approaches without tolerating missing data: ST-KNN approach (Cheng et al., 2018), T-GCN approach (Zhao et al., 2020), ASTGCN approach (Guo et al., 2019) and GDGCN approach (Y. Xu et al., 2023).
- Forecasting approaches with tolerating missing data: LSTM-M approach (Tian et al., 2018), BTMF approach (Chen & Sun, 2022), SGMN approach (Cui et al., 2020) and D-TGNM approach (P. Wang, Zhang, Hu et al., 2023).

4.3.2. Comparison results of forecasting accuracy on complete data

In scenarios without missing data, we only need to use the Predictor to address the spatiotemporal forecasting task, as shown in Table 2. Overall, under these no-missing scenarios, the forecasting accuracy of the second-category approaches surpasses that of the first-category approaches. Notably, the Predictor not only outperforms the baselines but also achieves comparable forecasting accuracy in some cases. Additionally, the results indicate that the Predictor exhibits superior generalization ability compared to the baselines, consistently achieving better forecasting accuracy across all three spatiotemporal datasets.

4.3.3. Comparison results of forecasting accuracy on incomplete data

In scenarios with missing data, we utilize the Imputer-Predictor architecture to tackle the forecasting task, as shown in Tables 3 and 4. The findings reveal notable differences among the two categories of models in handling missing scenarios. Particularly noteworthy is the superior forecasting accuracy of second-category approaches over first-category approaches. The rationale behind this is that fourth-category approaches possess the capability to automatically handle missing data.

In comparison, the STDM approach surpasses ST-KNN, T-GCN, ASTGCN, GDGCN, BTMF, LSTM-M, and SGMN approaches in evaluation

Table 2
Evaluation metrics (RMSE/MAPE) of STDM and baselines under no-missing scenarios.

Model	Traffic		PM$_{2.5}$		Temperature	
	1-step	3-steps	1-step	3-steps	1-step	3-steps
ST-KNN	9.36/	10.09/	19.08/	23.13/	1.12/	1.33/
	29.64	30.82	37.12	46.75	2.91	3.42
T-GCN	11.69/	13.81/	21.11/	24.16/	1.08/	1.32/
	38.98	40.84	40.99	46.15	3.27	3.59
ASTGCN	6.58/	7.97/	13.09/	20.32/	0.85/	1.17/
	25.58	27.93	23.88	36.02	2.11	3.06
GDGCN	6.20/	7.09/	11.23/	17.65/	0.49/	0.70/
	22.68	24.13	20.97	32.19	1.24	1.76
BTMF	6.23/	7.36/	12.48/	18.64/	0.82/	1.09/
	23.01	24.65	22.16	33.91	1.99	2.79
LSTM-M	6.35/	7.46/	12.68/	18.89/	0.58/	0.72/
	23.59	25.64	22.34	34.18	1.43	1.81
SGMN	6.22/	7.31/	12.74/	19.15/	0.53/	0.73/
	23.09	25.18	22.85	35.88	1.32	1.84
D-TGNM	6.19/	7.13/	11.45/	18.08/	0.48/	0.68/
	22.91	24.04	21.76	33.14	1.23	1.78
STDM	6.19/	7.02/	10.97/	17.01/	0.46/	0.67/
	22.49	23.91	20.34	31.72	1.18	1.71

The data in each cell of this table is separated by a "/" character, where the first value represents the model's RMSE metric and the second value represents the model's MAPE metric.

metrics, closely approaching the D-TGNM approach. Furthermore, focusing on the 40 % block missing dataset, we conduct an additional analysis comparing the computational efficiency between the STDM approach and baselines, as depicted in Fig. 9. The outcomes underscore the superior forecasting accuracy and significant computational

Table 3

Evaluation metrics (RMSE/MAPE) of STDM and baselines with random missing data.

Model	MR	Traffic		PM$_{2.5}$		Temperature	
		1-step	3-steps	1-step	3-steps	1-step	3-steps
ST-KNN	20%	19.40/37.51	19.68/38.36	29.34/44.70	31.64/51.97	7.94/20.66	8.04/21.13
	40%	30.02/50.37	30.28/51.37	42.72/58.74	44.14/63.78	13.01/38.97	13.17/39.54
T-GCN	20%	13.20/34.70	14.72/37.36	18.83/37.10	24.48/44.90	2.12/5.76	3.31/9.58
	40%	14.11/35.70	15.50/39.08	19.68/39.01	25.06/49.53	2.21/6.03	3.32/9.60
ASTGCN	20%	10.34/31.66	11.93/33.67	16.68/28.62	22.27/46.02	1.33/3.54	1.71/4.81
	40%	11.29/32.87	12.48/35.48	17.64/32.03	23.42/46.39	1.50/4.34	1.84/5.16
GDGCN	20%	8.35/29.17	9.04/31.02	14.17/25.04	20.91/45.35	1.03/2.81	1.31/3.45
	40%	9.43/31.64	10.68/35.91	15.34/29.68	21.09/46.35	1.05/2.85	1.38/3.78
BTMF	20%	7.28/26.00	8.41/28.43	12.86/24.03	19.01/35.65	0.84/2.11	1.03/2.79
	40%	7.89/27.61	8.92/30.19	14.22/26.54	20.65/38.98	0.91/2.62	1.18/3.27
LSTM-M	20%	7.43/24.91	8.53/27.79	14.87/26.88	20.68/39.61	0.54/1.86	0.95/2.35
	40%	8.06/28.42	9.08/29.63	17.52/32.01	22.20/44.46	0.63/2.23	1.15/2.88
SGMN	20%	7.31/26.35	8.34/28.63	14.62/25.32	19.31/38.32	0.55/1.43	0.84/2.23
	40%	7.86/27.63	8.93/31.97	15.56/27.10	21.36/41.65	0.58/1.47	0.86/2.26
D-TGNM	20%	7.16/24.43	8.16/27.63	12.76/23.36	18.97/36.15	0.54/1.40	0.78/2.09
	40%	7.83/26.98	8.86/30.39	14.10/25.87	20.34/39.13	0.59/1.49	0.81/2.21
STDM	**20%**	**6.96/24.37**	**7.99/27.69**	**12.21/22.35**	**18.31/34.06**	**0.52/1.34**	**0.74/1.91**
	40%	**7.71/26.72**	**8.62/29.22**	**13.56/25.05**	**19.19/36.76**	**0.55/1.44**	**0.76/1.95**

The data in each cell of this table is separated by a "/" character, where the first value represents the model's RMSE metric and the second value represents the model's MAPE metric.

efficiency advantage of the STDM approach over baselines.

4.4. Influence of various components on forecasting performance

Taking the 3-steps forecasting under the missing scenario as an example, this subsection analyzed the influence of various components on forecasting results, as shown in Table 5. Among them, the TDC corresponds to Eq. (7), the STD represents the integration of GDC operator into TDC operator, and the TDCM represents the integration of MDH component into TDC. Additionally, (R) denotes the random missing dataset, and (B) denotes the block missing dataset. The results indicate that the STD approach outperforms the TDC approach, demonstrating the importance of incorporating a graph attention operator to improve forecasting accuracy. Furthmore, the TDCM approach surpasses the TDC approach, highlighting the necessity of including a missing data handling component.

4.5. Influence of loss function on forecasting performance
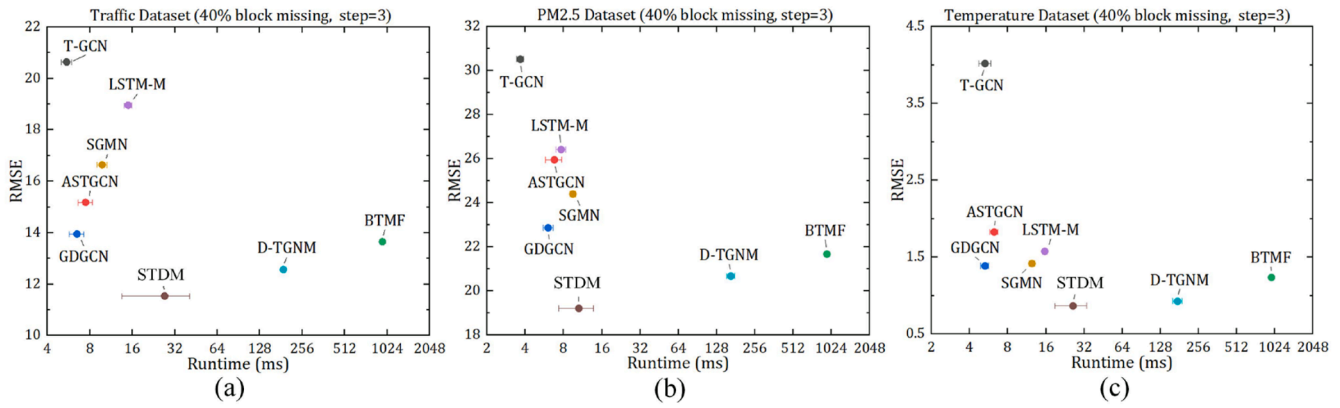
As mentioned earlier, we design the STDM approach as a separable architecture and define a corresponding loss function. Therefore, we analyzed the effect of the loss function on the STDM approach, as shown in Table 6. Among them, the STDM -NIL denotes the STDM approach without loss of Imputer, (R) represents the dataset with random missing pattern, and (B) represents the dataset with block missing pattern. The results indicate that the STDM approach achieves higher forecasting accuracy than the STDM-NIL approach under missing scenarios, demonstrating the effectiveness of the loss function design.

4.6. Qualitative analysis

4.6.1. Qualitative analysis of forecasting accuracy

Using datasets with 40 % block-missing and no missing data as examples, we visualize the difference between the forecasting values and ground truth in temporal dimension, as shown in Fig. 9. In general, the STDM approach demonstrates the ability to accurately predict data trends in the temporal dimension. The results confirm the robust forecasting performance of the STDM approach regardless of missing or non-missing scenarios. In addition, two situations can cause the forecasting accuracy to become less accurate under both missing and non-missing scenarios. One is that the trend of the spatiotemporal data changes abruptly over a short period, such as the blue areas in Fig. 10(a)–(c). Another is that there is missing data in the spatiotemporal data, such as

Table 4

Evaluation metrics (RMSE/MAPE) of STDM and baselines with block missing data.

Model	MR	Traffic		PM$_{2.5}$		Temperature	
		1-step	3-steps	1-step	3-steps	1-step	3-steps
ST-KNN	20 %	32.51/64.32	35.85/65.17	44.38/65.86	45.83/59.37	11.86/29.20	12.87/32.13
	40 %	36.47/79.52	38.30/76.27	49.71/65.56	49.55/69.16	14.30/38.48	14.83/39.94
T-GCN	20 %	17.24/36.70	19.91/41.91	19.31/39.82	29.09/49.89	2.61/7.78	3.53/10.19
	40 %	17.90/37.24	20.64/43.26	20.33/44.43	30.52/54.41	2.69/7.85	4.02/11.61
ASTGCN	20 %	11.45/32.14	12.13/34.98	17.39/31.07	24.62/47.59	1.40/3.93	1.79/5.05
	40 %	14.98/46.47	15.18/48.94	18.93/36.78	25.96/48.01	1.57/4.59	1.83/5.24
GDGCN	20 %	8.91/29.48	10.75/36.78	14.39/27.17	22.29/45.43	1.08/2.94	1.35/3.52
	40 %	12.97/38.47	13.95/40.86	15.95/30.95	22.88/45.48	1.12/3.27	1.39/3.69
BTMF	20 %	9.34/29.34	10.96/32.14	14.19/27.48	20.32/38.31	0.89/2.23	1.18/3.38
	40 %	12.68/34.98	13.65/37.69	16.67/33.08	21.69/41.32	0.97/2.81	1.24/3.54
LSTM-M	20 %	13.66/37.76	15.00/46.25	18.91/39.29	23.75/51.89	0.69/2.45	1.32/3.00
	40 %	18.17/51.35	18.96/67.43	23.18/51.46	26.43/61.21	0.92/3.27	1.58/3.81
SGMN	20 %	10.32/31.92	13.36/36.71	17.40/28.42	22.65/49.63	0.62/1.61	1.29/3.97
	40 %	14.16/36.67	16.64/49.98	22.32/45.31	26.41/56.18	0.86/2.19	1.42/4.03
D-TGNM	20 %	8.02/27.34	10.32/31.36	13.53/26.97	19.32/37.63	0.57/1.49	0.81/2.29
	40 %	12.04/33.64	12.57/36.79	15.91/32.07	20.69/41.35	0.70/1.96	0.93/2.52
STDM	**20 %**	**7.71/27.72**	**8.63/29.22**	**13.03/25.20**	**18.65/36.23**	**0.54/1.43**	**0.77/2.01**
	40 %	**11.07/31.31**	**11.54/34.19**	**14.60/30.12**	**19.23/40.35**	**0.66/1.63**	**0.87/2.18**

The data in each cell of this table is separated by a "/" character, where the first value represents the model's RMSE metric and the second value represents the model's MAPE metric.

Fig. 9. Forecasting accuracy vs. forward propagation runtime under 40 % block missing: (a) traffic dataset, (b) PM$_{2.5}$ dataset, and (c) temperature dataset.

Table 5
Evaluation metrics (RMSE/MAPE) of STDM and various components.

Model	Traffic		PM$_{2.5}$		Temperature	
	MR:20 %	MR:40 %	MR:20 %	MR:40 %	MR:20 %	MR:40 %
TDC (R)	10.86/ 87.27	12.08/ 107.6	26.61/ 58.83	27.98/ 66.49	2.22/ 6.45	2.50/ 7.20
TDC (B)	19.59/ 101.4	24.83/ 125.6	29.06/ 77.46	30.93/ 83.13	2.31/ 6.66	2.54/ 7.25
STD (R)	9.32/ 33.08	10.67/ 34.68	22.19/ 44.44	23.01/ 45.94	1.30/ 3.52	1.32/ 3.57
STD (B)	10.94/ 35.91	13.88/ 38.17	25.91/ 47.66	26.28/ 48.51	1.34/ 3.65	1.43/ 3.86
TDCM (R)	9.92/ 48.02	10.56/ 49.49	20.07/ 40.82	22.06/ 44.13	1.33/ 3.62	1.52/ 4.09
TDCM (B)	16.16/ 64.85	19.60/ 74.47	28.61/ 67.51	28.26/ 65.47	1.78/ 4.58	2.01/ 5.29
STDM (R)	**7.99/ 27.69**	**8.62/ 29.22**	**18.31/ 34.06**	**19.19/ 36.76**	**0.74/ 1.91**	**0.76/ 1.95**
STDM (B)	**8.63/ 29.22**	**11.54/ 34.19**	**18.65/ 36.23**	**19.23/ 40.35**	**0.77/ 2.01**	**0.87/ 2.18**

The data in each cell of this table is separated by a "/" character, where the first value represents the model's RMSE metric and the second value represents the model's MAPE metric.

Table 6
Evaluation metrics (RMSE/MAPE) of STDM and STDM-NIL approaches.

Dataset	STDM-NIL		STDM	
	MR: 20 %	MR: 40 %	MR: 20 %	MR: 40 %
Traffic Dataset (R)	8.14/29.31	9.40/31.68	7.99/27.69	8.62/29.22
Traffic Dataset (B)	11.14/ 33.28	13.02/ 36.43	8.63/29.22	11.54/ 34.19
PM$_{2.5}$ Dataset (R)	18.71/ 34.98	20.04/ 38.87	18.31/ 34.06	19.19/ 36.76
PM$_{2.5}$ Dataset (B)	19.62/ 39.27	22.36/ 45.35	18.65/ 36.23	19.23/ 40.35
Temperature Dataset (R)	0.77/1.98	0.81/2.18	0.74/1.91	0.76/1.95
Temperature Dataset (B)	0.87/2.24	1.01/2.87	0.77/2.01	0.87/2.18

The data in each cell of this table is separated by a "/" character, where the first value represents the model's RMSE metric and the second value represents the model's MAPE metric.

the blue areas in Fig. 10(d)–(f). Both scenarios mentioned above disrupt the spatiotemporal patterns in the data, thereby increasing the difficulty in model forecasting.

Fig. 11 provides further visualization of the disparity between the forecasting values and truth values in spatial dimension. The results indicate the STDM approach accurately predicts data changes in the spatial dimension, regardless of missing or non-missing scenarios. Similarly, like forecasting errors in the temporal dimension, fluctuations in spatiotemporal data can also affect forecasting accuracy in the spatial dimension. To substantiate this conclusion, we calculated the correlation coefficient between data fluctuations and forecasting difficulty. The results show a significant positive correlation between observed variance and forecasting accuracy across all three datasets. This further underscores the impact of data fluctuations on forecasting accuracy.

Overall, the STDM approach has excellent forecasting accuracy under both missing and non-missing scenarios, proving that the STDM approach can be applied to spatiotemporal forecasting tasks in various scenarios.

### 4.6.2. Qualitative analysis of automatically handing missing data

This subsection delves into the approach's capability to automatically handle missing data, as depicted in Fig. 12. Among them, the red line shows the relationship between u^t and β^t, and the blue line shows the relationship between u^t and $1 - \beta^t$. We found that noteworthy distinctions in the approaches to handling missing data among the three spatiotemporal datasets, despite employing a consistent forward propagation process. For instance, when the missing type leans towards random, the traffic dataset and PM$_{2.5}$ dataset leverage temporal nearest observations to estimate missing data, whereas the temperature dataset relies on spatial akin observations. Conversely, when the missing type tilts towards blocky, the traffic dataset and PM$_{2.5}$ dataset resort to spatial akin observations, whereas the temperature dataset turns to temporal nearest observations. These differences arise from the automatic data processing strategies inherent in the STDM approach. This autonomous adaptability elucidates why the STDM approach demonstrates superior performance in scenarios involving missing data.

4.6.3. Qualitative analysis of integrated graph attention

During the approach construction process, graph attention is used to improve the forecasting accuracy. The reason is that graph attention has the ability to capture geospatial heterogeneity. In this subsection, we analyze the ability of graph attention to capture geospatial heterogeneity. As is shown in Fig. 13, there is a significant difference in the weight of influence of neighboring spatial objects on the target spatial object under a single time window. At the same time, we find that the neighboring spatial objects closer to the target spatial objects have greater influence weight on the target spatial objects. For example, the influence weight of neighboring spatial objects on the target spatial object decreases sequentially centered on the target spatial object and outward, aligning with our common sense. In addition, Fig. 14 further visualizes the geospatial relationships between spatial objects over time. The findings reveal significant variations in the influence weights of neighboring spatial objects on target spatial objects across multiple time windows. These results underscore how the proposed STDM effectively
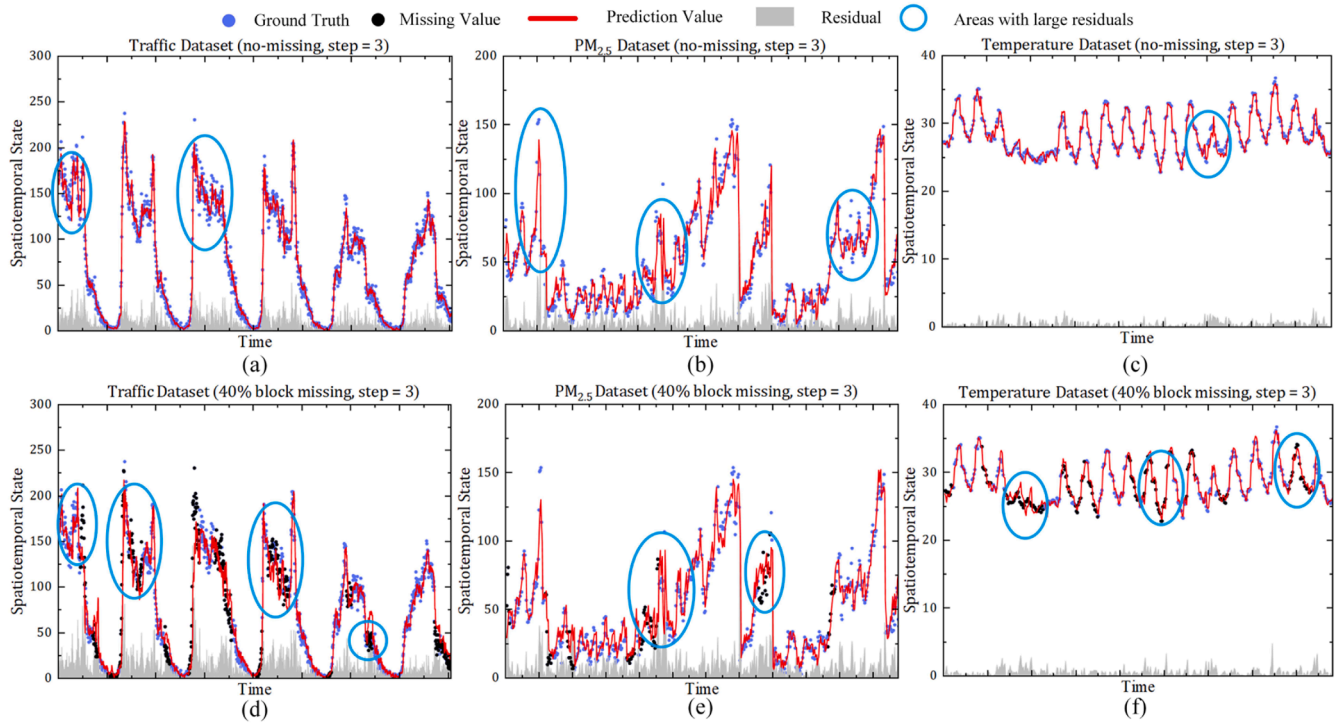
Fig. 10. Forecasting performance in time dimension: (a) traffic dataset without missing data, (b) PM$_{2.5}$ dataset without missing data, (c) temperature dataset without missing data, (d) traffic dataset with 40 % block missing data, (e) PM$_{2.5}$ dataset with 40 % block missing data, and (f) temperature dataset with 40 % block missing data.
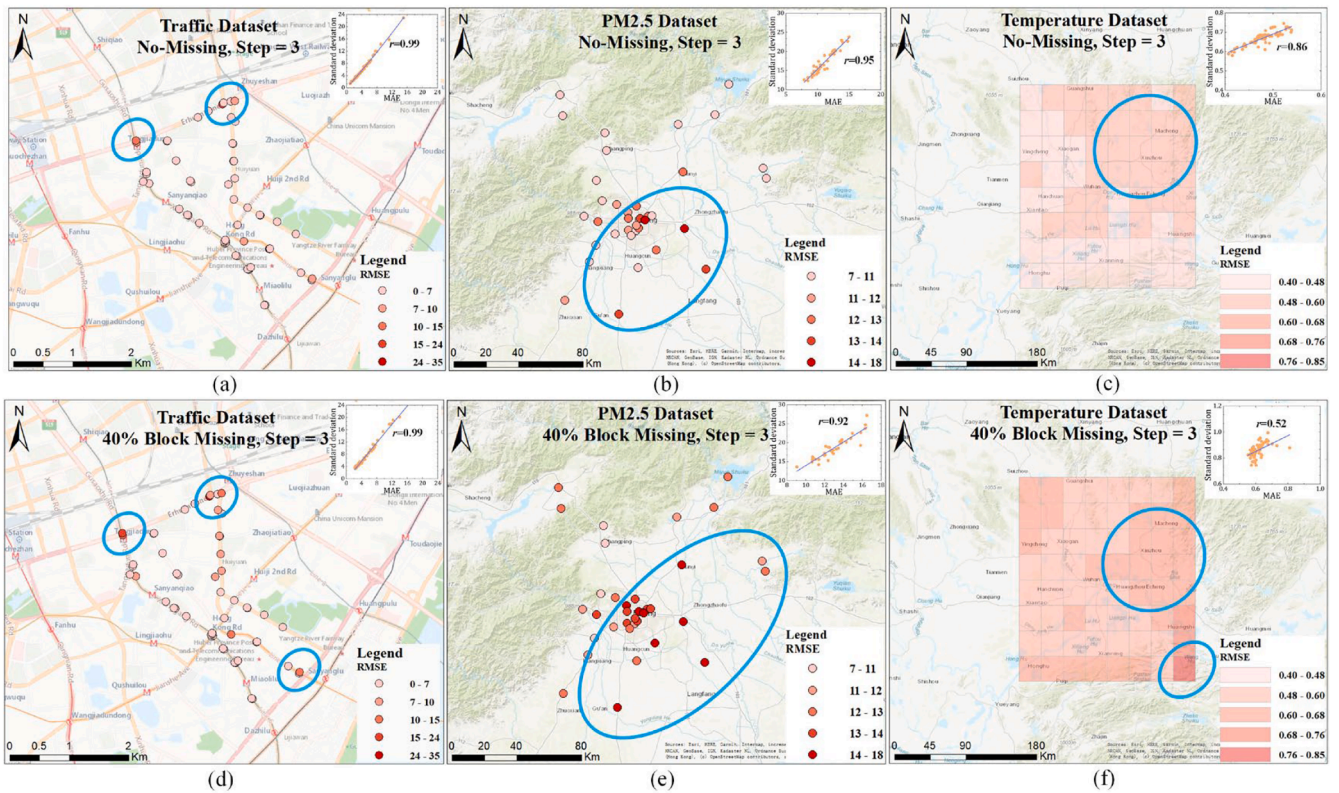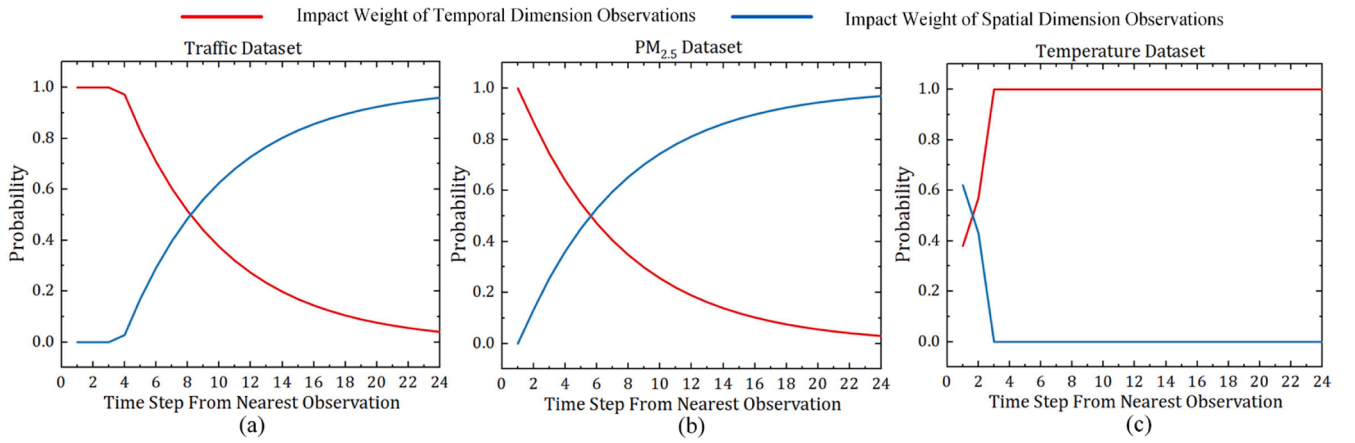

Fig. 11. Forecasting performance in spatial dimension: (a) traffic dataset without missing data, (b) PM$_{2.5}$ dataset without missing data, (c) temperature dataset without missing data, (d) traffic dataset with 40 % block missing data, (e) PM$_{2.5}$ dataset with 40 % block missing data, and (f) temperature dataset with 40 % block missing data.

Fig. 12. Automatic learning of missing data processing strategies:(a) traffic dataset, (b)PM$_{2.5}$ dataset, and (c) temperature dataset.
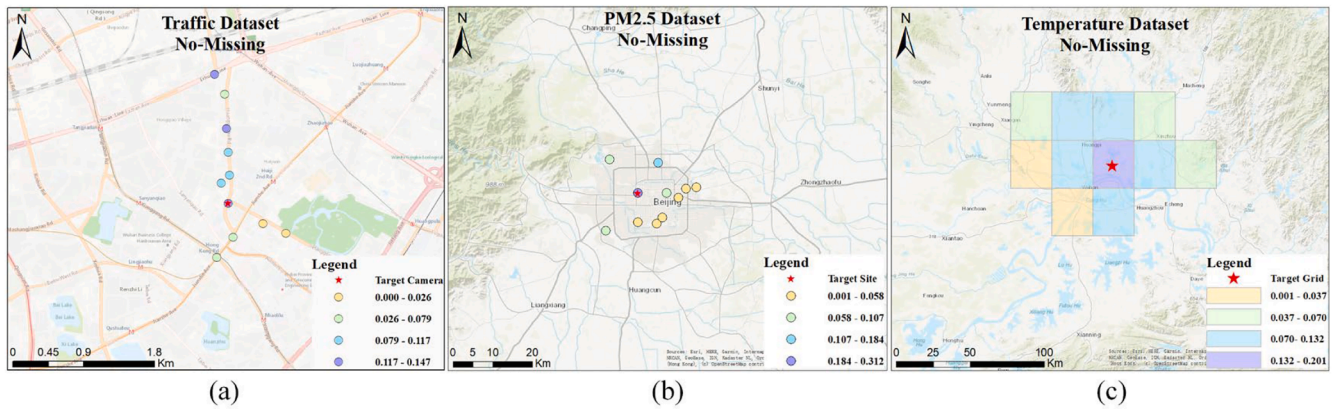

Fig. 13. llustration of geospatial heterogeneity under a specific time window:(a) traffic dataset, (b)PM$_{2.5}$ dataset, and (c) temperature dataset.
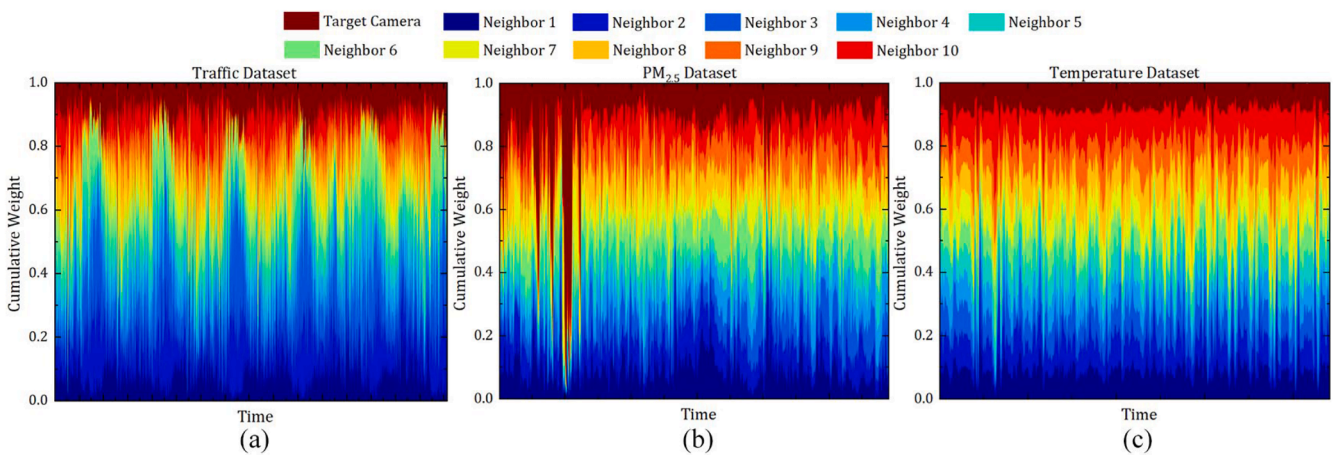

Fig. 14. llustration of geospatial heterogeneity under multiple time windows:(a) traffic dataset, (b)PM$_{2.5}$ dataset, and (c) temperature dataset.

captures geospatial heterogeneity, elucidating why it outperforms baseline models in non-missing scenarios.

5. Discussion

This study proposed the STDM approach for fast forecasting of urban sensor states under missing scenes. Compared to forecasting approaches that cannot tolerate missing data (Cheng et al., 2018; Guo et al., 2019; Y. Xu et al., 2023; Zhao et al., 2020), the proposed STDM approach

demonstrates strong robustness in complex missing data scenarios, achieving forecasting accuracy improvements of over three times. Compared to forecasting approaches that tolerate missing data (Chen & Sun, 2022; Cui et al., 2020; Tian et al., 2018; P. Wang, Zhang, Hu et al., 2023), the proposed STDM approach demonstrates a fast inference speed in complex missing data scenarios, surpassing the baselines by a factor of 10 or more. In addition, the proposed STDM approach offers visualization tools to help users understand the automated processing strategies for missing data. Finally, the STDM approach features a

modular Imputer-Predictor architecture, enabling users to freely select either the Predictor module or the full Imputer-Predictor setup based on the presence or absence of missing data in real-world scenarios. In summary, we can confidently assert that the STDM approach represents an advanced forecasting method.

There were also some limitations in this study. For example, there are many high missing rate scenarios in practical applications. However, the STDM approach was only verified in low missing rate scenarios. Given the above problems, future work will apply the STDM approach to scenarios with higher missing rates. In addition, the STDM approach is a generalized spatiotemporal forecasting model, and future work will apply the STDM approach to more spatiotemporal datasets.

6. Conclusion

In this study, we established a new forecasting approach, i.e., the STDM approach. In experimental results and analysis section, the forecasting performance of the STDM approach was evaluated using three real spatiotemporal datasets: traffic, $PM_{2.5}$, and temperature datasets. Initially, we compared the STDM approach with eight baselines and demonstrated its forecasting accuracy and inference speed advantages under four missing scenarios. Additionally, we conducted an analysis of the impact of various components and loss functions on forecasting performance, thereby validating the rationale behind the design of different components and loss functions in our approach. Finally, we explained the reasons behind the performance of the STDM approach through visualization.

Data and codes availability statement

The data and codes are available on https://doi.org/10.6084/m9.figshare.24289456.

CRediT authorship contribution statement

Peixiao Wang: Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Hengcai Zhang:** Writing – review & editing, Methodology. **Shifen Cheng:** Writing – review & editing, Conceptualization. **Tong Zhang:** Writing – review & editing, Funding acquisition. **Feng Lu:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare no conflicts of interest.

Funding

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scs.2024.106044.

References

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports, 8*(1), 6085. https://doi.org/10.1038/s41598-018-24271-9

Chen, X., & Sun, L. (2022). Bayesian temporal factorization for multidimensional time series prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*, 4659–4673. https://doi.org/10.1109/TPAMI.2021.3066551

Cheng, S., Lu, F., Peng, P., & Wu, S. (2018). Short-term traffic forecasting: An adaptive ST-KNN model that considers spatial heterogeneity. *Computers, Environment and Urban Systems, 71*, 186–198. https://doi.org/10.1016/j.compenvurbsys.2018.05.009

Cui, Z., Lin, L., Pu, Z., & Wang, Y. (2020). Graph Markov network for traffic forecasting with missing data. *Transportation Research Part C: Emerging Technologies, 117*, Article 102671. https://doi.org/10.1016/j.trc.2020.102671

Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., & He, S. (2019). Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In *, 33. Proceedings of the AAAI conference on artificial intelligence.* https://doi.org/10.1609/aaai.v33i01.3301890. Article 01.

Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies, 108*, 12–28. https://doi.org/10.1016/j.trc.2019.09.008

Du, H., Zhou, Y., Ma, Y., & Wang, S. (2021). Astrologer: Exploiting graph neural Hawkes process for event propagation prediction with spatio-temporal characteristics. *Knowledge-Based Systems, 228*, Article 107247. https://doi.org/10.1016/j.knosys.2021.107247

Fang, M., Tang, L., Yang, X., Chen, Y., Li, C., & Li, Q. (2022). FTPG: A fine-grained traffic prediction method with graph attention network using big trace data. *IEEE Transactions on Intelligent Transportation Systems, 23*(6), 5163–5175. https://doi.org/10.1109/TITS.2021.3049264

Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*, 922–929. https://doi.org/10.1609/aaai.v33i01.3301922

Jiang, F., Ma, J., & Li, Z. (2022). Pedestrian volume prediction with high spatiotemporal granularity in urban areas by the enhanced learning model. *Sustainable Cities and Society, 79*, Article 103653. https://doi.org/10.1016/j.scs.2021.103653

Khaled, A., Elsir, A. M. T., Wang, P., Shen, Y., & Zhang, Q. (2024). A graph-based approach for traffic prediction using similarity and causal relations between nodes. *Knowledge-Based Systems, 296*, Article 111913. https://doi.org/10.1016/j.knosys.2024.111913

Li, G., Zhong, S., Deng, X., Xiang, L., Chan, S.-H. G., Li, R., Liu, Y., Zhang, M., Hung, C.-C., & Peng, W.-C. (2023). A Lightweight and Accurate Spatial-Temporal Transformer for Traffic Forecasting. *IEEE Transactions on Knowledge and Data Engineering, 35*(11), 10967–10980. https://doi.org/10.1109/TKDE.2022.3233086

Li, H., Li, M., Lin, X., He, F., & Wang, Y. (2020). A spatiotemporal approach for traffic data imputation with complicated missing patterns. *Transportation Research Part C: Emerging Technologies, 119*, Article 102730. https://doi.org/10.1016/j.trc.2020.102730

Liu, Y., Wang, X., Hou, W., Liu, H., & Wang, J. (2022). A novel hybrid model combining a fuzzy inference system and a deep learning method for short-term traffic flow prediction. *Knowledge-Based Systems, 255*, Article 109760. https://doi.org/10.1016/j.knosys.2022.109760

Liu, Z., Zhang, S., Shao, X., & Wu, Z. (2023). Accurate and efficient urban wind prediction at city-scale with memory-scalable graph neural network. *Sustainable Cities and Society, 99*, Article 104935. https://doi.org/10.1016/j.scs.2023.104935

Mei, H., Li, J., Liang, Z., Zheng, G., Shi, B., & Wei, H. (2023). *Uncertainty-aware traffic prediction under missing data* (arXiv:2309.06800). arXiv. https://doi.org/10.48550/arXiv.2309.06800.

Palanisamy, P. A., Zawadzka, J., Jain, K., Bonafoni, S., & Tiwari, A. (2024). Assessing diurnal land surface temperature variations across landcover and local climate zones: Implications for urban planning and mitigation strategies on socio-economic factors. *Sustainable Cities and Society, 116*, Article 105880. https://doi.org/10.1016/j.scs.2024.105880

Qu, Y., Rong, J., Li, Z., & Chen, K. (2023). ST-A-PGCL: Spatiotemporal adaptive periodical graph contrastive learning for traffic prediction under real scenarios. *Knowledge-Based Systems, 272*, Article 110591. https://doi.org/10.1016/j.knosys.2023.110591

Rabie, R., Asghari, M., Nosrati, H., Emami Niri, M., & Karimi, S. (2024). Spatially resolved air quality index prediction in megacities with a CNN-Bi-LSTM hybrid framework. *Sustainable Cities and Society, 109*, Article 105537. https://doi.org/10.1016/j.scs.2024.105537

Soudeep, S., Lailun Nahar Aurthy, Most., Jim, J. R., Mridha, M. F., & Kabir, M. M. (2024). Enhancing road traffic flow in sustainable cities through transformer models: Advancements and challenges. *Sustainable Cities and Society, 116*, Article 105882. https://doi.org/10.1016/j.scs.2024.105882

Tian, Y., Zhang, K., Li, J., Lin, X., & Yang, B. (2018). LSTM-based traffic flow prediction with missing data. *Neurocomputing, 318*, 297–305. https://doi.org/10.1016/j.neucom.2018.08.067

Wang, P., Zhang, H., Cheng, S., Zhang, T., Lu, F., & Wu, S. (2024). A lightweight spatiotemporal graph dilated convolutional network for urban sensor state prediction. *Sustainable Cities and Society, 101*, Article 105105. https://doi.org/10.1016/j.scs.2023.105105

Wang, P., Zhang, T., Nie, S., Yang, J., & Wang, T. (2023). A causal graph convolutional network considering missing values for spatiotemporalprediction. *Acta Geodaetica et Cartographica Sinica, 52*(5), 818–830. https://doi.org/10.11947/j.AGCS.2023.20220021 (in Chinese).

Wang, P., Zhang, Y., Hu, T., & Zhang, T. (2023). Urban traffic flow prediction: A dynamic temporal graph network considering missing values. *International Journal of*

Geographical Information Science, 37(4), 885–912. https://doi.org/10.1080/13658816.2022.2146120

Wang, Z., Ran, H., Ren, J., & Sun, M. (2024). PWDformer: Deformable transformer for long-term series forecasting. *Pattern Recognition, 147*, Article 110118. https://doi.org/10.1016/j.patcog.2023.110118

Xu, L., Chen, N., Chen, Z., Zhang, C., & Yu, H. (2021). Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. *Earth-Science Reviews, 222*, Article 103828. https://doi.org/10.1016/j.earscirev.2021.103828

Xu, M., Dai, W., Liu, C., Gao, X., Lin, W., Qi, G.-J., & Xiong, H. (2021). Spatial-Temporal Transformer Networks for Traffic Flow Forecasting. *arXiv:2001.02908 [Cs, Eess]*. http://arxiv.org/abs/2001.02908.

Xu, Y., Han, L., Zhu, T., Sun, L., Du, B., & Lv, W. (2023). Generic dynamic graph convolutional network for traffic flow forecasting. *Information Fusion.* , Article 101946. https://doi.org/10.1016/j.inffus.2023.101946

Xu, Z., Lv, Z., Chu, B., & Li, J. (2023). Fast autoregressive tensor decomposition for online real-time traffic flow prediction. *Knowledge-Based Systems, 282*, Article 111125. https://doi.org/10.1016/j.knosys.2023.111125

Yu, H.-F., Rao, N., & Dhillon, I. S. (2016). Temporal regularized matrix factorization for high-dimensional time series prediction. In *30th Conference on neural information processing systems (NIPS 2016)* (p. 15).

Yu, J. J. Q (2021). Citywide traffic speed prediction: A geometric deep learning approach. *Knowledge-Based Systems, 212*, Article 106592. https://doi.org/10.1016/j.knosys.2020.106592

Zhang, B., Cheng, S., Zhao, Y., & Lu, F. (2023). Inferring intercity freeway truck volume from the perspective of the potential destination city attractiveness. *Sustainable Cities and Society, 98*, Article 104834. https://doi.org/10.1016/j.scs.2023.104834

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. *Proceedings of the AAAI Conference on Artificial Intelligence, 31*(1).

Zhang, M., Li, Z., Si, H., Cheng, L., Zhou, X., & Wang, B. (2023). Urban travel time and residential location choice: The impacts of traffic congestion. *Sustainable Cities and Society, 99*, Article 104975. https://doi.org/10.1016/j.scs.2023.104975

Zhang, Y., Cheng, T., Ren, Y., & Xie, K. (2020). A novel residual graph convolution deep learning model for short-term network-based traffic forecasting. *International Journal of Geographical Information Science, 34*(5), 969–995. https://doi.org/10.1080/13658816.2019.1697879

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020). T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems, 21*(9), 3848–3858. https://doi.org/10.1109/TITS.2019.2935152

Zhao, L., Zhou, Y., Lu, H., & Fujita, H. (2019). Parallel computing method of deep belief networks and its application to traffic flow prediction. *Knowledge-Based Systems, 163*, 972–987. https://doi.org/10.1016/j.knosys.2018.10.025

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology, 5*(3), 1–55. https://doi.org/10.1145/2629592